

Statistical Methods in Epidemiology

Lab 8–9

Propensity score methods

Στόχος:

Η κατανόηση των βασικών εννοιών και τεχνικών της ανάλυσης propensity score, με έμφαση στη διαδικασία εξισορρόπησης συγχυτικών παραγόντων σε μη τυχαιοποιημένες μελέτες. Στο παρόν εργαστήριο θα εφαρμοστούν οι μέθοδοι:

- Propensity score estimation
- Έλεγχος ισορροπίας (balance checking)
- Stratification
- IPTW (Inverse Probability of Treatment Weighting)

Dataset: **psadata.dta**

1. Εισαγωγή: Propensity Scores και Confounding

Σε μη τυχαιοποιημένες μελέτες, η εκτιμώμενη επίδραση μιας θεραπείας μπορεί να είναι μεροληπτική εξαιτίας συγχυτικών παραγόντων που επηρεάζουν τόσο τη θεραπεία όσο και το αποτέλεσμα.

Το propensity score είναι η πιθανότητα να λάβει κάποιος θεραπεία, δεδομένων των συγχυτικών παραγόντων:

$$e(X) = P(T = 1 | X)$$

όπου

- $T = 1$ αν έλαβε θεραπεία, 0 αν όχι
- X = σύνολο συγχυτικών παραγόντων

Για άτομα με ίδιο propensity score, η κατανομή των X είναι ίδια μεταξύ treated και untreated.

2. Φόρτωση δεδομένων και έλεγχος αρχικής ανισορροπίας

1. Ανοίξτε το αρχείο **psadata.dta**. Ποιες μεταβλητές περιλαμβάνει το dataset;
2. Ελέγξτε αν η μέση τιμή του outcome διαφέρει μεταξύ treated και untreated.
3. Εξετάστε αν η ανάθεση της θεραπείας εξαρτάται από τα χαρακτηριστικά των ασθενών. Μπορείτε να θεωρήσετε ότι η θεραπεία έχει χορηγηθεί τυχαία;
4. Συγκρίνετε τους μέσους όρους και τις τυπικές αποκλίσεις των x_1 , x_2 ανά ομάδα θεραπείας.
5. Γιατί η απλή διαφορά μέσων x_1 και x_2 μεταξύ treated και untreated δεν είναι άμεσα συγκρίσιμη; Ποιος δείκτης είναι καταλληλότερος για την αξιολόγηση της ανισορροπίας;
6. Υπολογίστε τις standardized mean differences (SMDs) των x_1 και x_2 μεταξύ treated και untreated και σχολιάστε το μέγεθος της ανισορροπίας.

```
. pbalchk t x1 x2
```

3. Εκτίμηση του Propensity Score

7. Εκτιμήστε την πιθανότητα θεραπείας $Pr(T=1|X)$ με λογιστική παλινδρόμηση και αποθηκεύστε την ως νέα μεταβλητή (propensity).

```
. logistic t x1 x2
. predict propensity, pr
```

8. Απεικονίστε τις κατανομές (kernel density) του propensity score ξεχωριστά για treated και untreated και σχολιάστε την επικάλυψη μεταξύ των δύο κατανομών.

```
. graph twoway ///
(kdensity propensity if t==0, legend(label(1 "Untreated"))) ///
(kdensity propensity if t==1, legend(label(2 "Treated"))), ///
legend(position(1) ring(0))
```

9. Υπολογίστε τη γραμμική πρόβλεψη (logit του propensity score) από το λογιστικό μοντέλο και απεικονίστε τις κατανομές της για treated και untreated.

```
. predict lp, xb
. graph twoway ///
(kdensity lp if t==0, legend(label(1 "Untreated"))) ///
(kdensity lp if t==1, legend(label(2 "Treated"))), ///
legend(position(1) ring(0))
```

10. Εξετάστε, με κατάλληλους όρους στο λογιστικό μοντέλο, αν υπάρχουν σημαντικές αλληλεπιδράσεις ή μη γραμμικότητα στη σχέση των x1 και x2 με την πιθανότητα θεραπείας.

```
. foreach v of varlist x1 x2{
    foreach v2 of varlist x1 x2{
        cap drop temp
        gen temp=`v'*`v2'
        logit t x1 x2 temp
        di "Testing `v'*`v2'"
    }
}
```

1^η Μέθοδος (Στρωματοποίηση)

11. Δημιουργήστε 5 strata χρησιμοποιώντας quintiles του propensity score και παρουσιάστε την κατανομή treated και untreated μέσα σε κάθε stratum.

```
. xtile pscats = propensity, n(5)
. tab pscats t, row
```

12. Ελέγξτε την ισορροπία των x1 και x2 εντός των πέντε strata χρησιμοποιώντας standardized mean differences.

```
. pbalchk t x1 x2, strata(pscats)
```

13. Δημιουργήστε 10 strata (deciles) του propensity score και παρουσιάστε εκ νέου την κατανομή treated και untreated ανά stratum.

```
. xtile pscats10 = propensity, n(10)
. tab pscats10 t, row
```

14. Επαναλάβετε τον έλεγχο ισορροπίας για τα deciles και συγκρίνετε την ισορροπία των covariates σε σχέση με τα quintiles.

```
. pbalchk t x1 x2, strata(pscats10)
```

15. Εκτιμήστε την επίδραση της θεραπείας χρησιμοποιώντας παλινδρόμηση του outcome στο t και στους συγχυτικούς παράγοντες x1 και x2.

```
. regression y t x1 x2
```

16. Εκτιμήστε την επίδραση της θεραπείας εντός των 5 strata εφαρμόζοντας παλινδρόμηση του outcome στο t και σε δείκτες stratum.

```
. regression y t i.pscats
```

16. Υπολογίστε την επίδραση της θεραπείας με stratified regression με 10 strata

```
. regression y t i.pscats10
```

17. Εκτιμήστε την επίδραση της θεραπείας εντός των 10 strata με παρόμοιο μοντέλο παλινδρόμησης και συγκρίνετέ το με την εκτίμηση από τα 5 strata.

```
. regression y t i.pscats10
```

2^η Μέθοδος (Στρωματοποίηση)

18. Υπολογίστε τα inverse probability weights για τα άτομα που έλαβαν θεραπεία και για τα άτομα που δεν έλαβαν θεραπεία, χρησιμοποιώντας τις εκτιμημένες πιθανότητες θεραπείας.

```
. gen ipw=1/propensity if t==1
. replace ipw=1/(1-propensity) if t==0
```

19. Εξετάστε την ισορροπία των x1 και x2 μετά την εφαρμογή των IPTW, χρησιμοποιώντας standardized mean differences.

```
. pbalchk t x1 x2, wt(ipw)
```

20. Παρουσιάστε την κατανομή των IPTW. Εντοπίστε τυχόν ακραίες τιμές και σχολιάστε την παρουσία ή μη αστάθειας στα βάρη.

```
. summ ipw
. histogram ipw, bin(40)
```

21. Εκτιμήστε την επίδραση της θεραπείας χρησιμοποιώντας γραμμική παλινδρόμηση με pweights ίσα με τα IPTW.

```
. reg y t [pweight = ipw]
```

22. Συγκρίνετε την εκτίμηση της θεραπευτικής επίδρασης που προκύπτει από τα IPTW με τις αντίστοιχες εκτιμήσεις από τα quintiles και τα deciles του propensity score.

3^η Μέθοδος (1:1 Nearest Neighbor)

23. Εκτελέστε 1:1 nearest-neighbor matching χωρίς αντικατάσταση, χρησιμοποιώντας το propensity score ως μεταβλητή αντιστοίχισης και ορίστε ως estimand το ATE.

```
. teffects psmatch (y) (t x1 x2), nneighbor(1) ate
```

24. Παρουσιάστε το μέγεθος του matched sample: πόσοι treated και πόσοι untreated περιλαμβάνονται μετά το matching;

```
. tebalance summarize
```

25. Ελέγξτε την ισορροπία των x1 και x2 στο matched sample, χρησιμοποιώντας standardized mean differences ή/και κατάλληλες εντολές balance diagnostics.

26. Συγκρίνετε την ισορροπία των x1 και x2 μετά το matching με την αντίστοιχη ισορροπία:

- (α) πριν από οποιαδήποτε προσαρμογή
- (β) μετά τη stratification (quintiles/deciles)
- (γ) μετά τα IPTW.

27. Εκτιμήστε την επίδραση της θεραπείας στο matched sample, όπως αυτή προκύπτει από το μοντέλο matching που χρησιμοποιήσατε.

28. Συγκρίνετε την εκτιμώμενη επίδραση της θεραπείας από το matching με τις αντίστοιχες εκτιμήσεις από:

- τη naïve ανάλυση,
- τη stratification,
- τα IPTW,

και σχολιάστε τις ομοιότητες και διαφορές.

4^η Μέθοδος (Covariate Adjustment Using the Propensity Score)

29. Εκτιμήστε την επίδραση της θεραπείας χρησιμοποιώντας γραμμική παλινδρόμηση του outcome στο t και στο propensity score ως επιπλέον covariate.

```
. reg y t propensity
```

30. Εξετάστε αν η σχέση του propensity score με το outcome είναι γραμμική. Χρησιμοποιήστε fractional polynomials ή κατάλληλους μη γραμμικούς όρους.

```
. fp <propensity>, replace:reg y t <propensity>
```

31. Εκτιμήστε εκ νέου την επίδραση της θεραπείας αφού προσαρμόσετε το μοντέλο συμπεριλαμβάνοντας την κατάλληλη (πιθανώς μη γραμμική) μορφή του propensity score.

```
. reg y t propensity_1 propensity_2
```

32. Συγκρίνετε την εκτίμηση της θεραπευτικής επίδρασης που προκύπτει από το covariate adjustment με εκείνες από:

- τη naïve ανάλυση,
- τη regression adjustment (με x1, x2),
- τη stratification,
- τα IPTW,
- το matching.

33. Σχολιάστε αν η εκτίμηση της θεραπευτικής επίδρασης μετά το covariate adjustment είναι πιο κοντά στον conditional ή στον marginal effect και εξηγήστε το γιατί.

II. Non-collapsibility of odds ratios

34. Φορτώστε το αρχείο `noncollaps.dta` και εκτελέστε λογιστική παλινδρόμηση της έκβασης στο `exposure`. Καταγράψτε την εκτίμηση της επίδρασης.

```
. use noncollaps.dta, clear  
. logit event exposure
```

35. Προσθέστε στο μοντέλο τη μεταβλητή `stratum` και εκτελέστε ξανά λογιστική παλινδρόμηση της έκβασης στο `exposure` και στο `stratum`. Συγκρίνετε την εκτίμηση της επίδρασης με την προηγούμενη.

```
. logit event exposure stratum
```

36. Εξετάστε αν το `stratum` πληροί τα κριτήρια για να θεωρηθεί confounder:

(α) σχετίζεται με το `exposure`;

```
. logit exposure stratum
```

(β) σχετίζεται με το `outcome`;

```
. logit event stratum
```

(γ) δεν βρίσκεται στην αιτιακή αλυσίδα μεταξύ `exposure` και `outcome`.

37. Εκτιμήστε αν υπάρχει στατιστική αλληλεπίδραση μεταξύ `exposure` και `stratum`, προσθέτοντας τον αντίστοιχο όρο στο μοντέλο.

```
. xi: logit event i.exposure*stratum
```

38. Εξηγήστε αν η διαφορά των εκτιμήσεων του `exposure` μεταξύ του μοντέλου χωρίς `stratum` και του μοντέλου με `stratum` οφείλεται σε confounding ή σε non-collapsibility.